# Statistics on Multivariate Normal Distributions: A Geometric Approach and its Application to Diffusion Tensor MRI

Christophe Lenglet, Mikaël Rousson, Rachid Deriche, Olivier Faugeras

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *Statistics on Multivariate Normal Distributions: A Geometric Approach and its Application to Diffusion Tensor MRI*

Christophe Lenglet — Mikaël Rousson — Rachid Deriche — Olivier Faugeras

**N° 5242**

June 2004

Thème BIO

Rapport de recherche

ISSN 0249-6399

# Statistics on Multivariate Normal Distributions: A Geometric Approach and its Application to Diffusion Tensor MRI

Christophe Lenglet* , Mikaël Rousson† , Rachid Deriche‡ , Olivier Faugeras§

Thème BIO — Systèmes biologiques
Projet Odyssée

Rapport de recherche n° 5242 — June 2004 — 25 pages

**Abstract:** This report is dedicated to the statistical analysis of the space of multivariate normal probability density functions. It relies on the differential geometrical properties of the underlying parameter space, endowed with a Riemannian metric, as well as on recent works that led to the generalization of the normal law on non-linear spaces. We will first proceed to the state of the art in section 1, while expressing some quantities related to the structure of the manifold of interest, and then focus on the derivation of closed-form expressions for the mean, covariance matrix, modes of variation and normal law between multivariate normal distributions in section 2. We will also address the derivation of accurate and efficient numerical schemes to estimate the proposed quantities. A major application of the present work is the statistical analysis of diffusion tensor Magnetic Resonance Imaging. We show promising results on synthetic and real data in section 3.

**Key-words:** multivariate normal distribution, symmetric positive-definite matrix, mean, covariance matrix, modes of variation, Fisher information, geodesics, geodesic distance, Riemann-Christoffel tensor, Ricci tensor, scalar curvature, sectional curvature, homogeneous space, Lie group, Riemannian geometry

* Christophe.Lenglet@sophia.inria.fr
† Mikaël.Rousson@sophia.inria.fr
‡ Rachid.Deriche@sophia.inria.fr
§ Olivier.Faugeras@sophia.inria.fr

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 —Télécopie : +33 4 92 38 77 65

# Statistiques sur les Distributions Normales Multivariées: Une Approche Géométrique et son Application à l'IRM du Tenseur de Diffusion

**Résumé :** Ce rapport aborde l'analyse statistique de l'espace des densités de probabilité normales multivariées. Il s'appuie sur les propriétés, issues de la géométrie différentielle, de l'espace des paramètres sous-jacents équipé d'une métrique Riemannienne, ainsi que sur de récents travaux ayant mené à la généralisation de la loi normale aux espaces non-linéaires. Nous commencerons par dresser un état de l'art en section 1, tout en exprimant certains quantités liées à la structure de la notre espace d'intérêt. Nous nous concentrerons ensuite sur la dérivation d'expressions explicites pour la moyenne, la matrice de covariance, les modes de variation et pour une loi normale entre distributions normales multivariées en section 2. Nous aborderons également le développement de schémas numériques efficaces pour estimer les quantités proposées. Une application majeure du présent travail réside dans l'analyse statistique des images IRM du tenseur de diffusion. Nous présentons d'encourageants résultats sur des données réelles et synthétiques en section 3.

**Mots-clés :** distribution normale multivariée, matrice symétrique définie-positive, moyenne, matrice de covariance, modes de variation, information de Fisher, géodésiques, distance géodésique, tenseur de Riemann-Christoffel, tenseur de Ricci, courbure scalaire, courbure sectionelle, espace homogène, groupe de Lie, géométrie Riemannienne

# Contents

**1 Geometry of the Space of Multivariate Normal Distributions 4**
1.1 Metrization of the Space of Probability Density Functions . . . . . . . . . . . 4
1.2 Information Geodesic Distance . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
1.3 Application to the Multivariate Normal Model . . . . . . . . . . . . . . . . . . 6
1.4 Structure of the Space of Multivariate Normal Distributions . . . . . . . . . . . 10

**2 Statistics on Multivariate Normal Distributions 13**
2.1 Intrinsic Mean . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
2.2 Intrinsic Covariance Matrix and Principal Modes . . . . . . . . . . . . . . . . . 16
2.3 A Normal Distribution on Multivariate Normal Distributions . . . . . . . . . . 17

**3 Numerical Experimentations 18**

**4 Conclusion 22**

**5 Appendix 1: Components of the Riemannian metric 23**

# 1 Geometry of the Space of Multivariate Normal Distributions

We review important notions that led to the metrization of probability spaces and to the definition of distance measures between parametric distributions under certain regularity conditions. Following the work by Rao [21], where a Riemannian metric was introduced in term of the elements of the information matrix, and a theorem by Jensen (1976, private communication in [2]), Atkinson and Mitchell obtained closed forms of the geodesic distance between elements of well-known families of distributions such as the fixed-mean multivariate normal model. In 1982, Burbea and Rao proposed a unified approach to the derivation of metrics in probability spaces. They introduced the notion of $\phi$-entropy functional whose Hessian in a direction of the tangent space of the parameter space is taken as the metric. We also follow [24], [3], [5] to introduce a Riemannian metric on the parameter space of the multivariate normal distributions, and the associated geodesic distance, Riemann-Christoffel tensor, Ricci tensor and Ricci scalar.

## 1.1 Metrization of the Space of Probability Density Functions

We start by a general characterization of the space of probability density functions, together with the characterization of its metric.
Let $L^1(\mathcal{X}, \mu)$ denote the space of integrable $\mu$-measurable real functions defined over the measurable space $\mathcal{X}$, e.g:

$$L^1(\mathcal{X}, \mu) = \{p : \|p\|_\mu = \int_{\mathcal{X}} |p(x)| d\mu(x) < \infty\}$$

We are interested in the convex subset $\mathcal{P}$ of $L^1_+$ such that:

$$L^1_+ = \{p \in L^1(\mathcal{X}, \mu) : p(x) \geq 0 \text{ for } \mu\text{-almost all } x \in \mathcal{X}\}$$

$$\mathcal{P} = \{p \in L^1_+ : \|p\|_\mu = 1\}$$

Let $\phi$ be a continuous real function on $I_\phi \subset \mathbb{R}^+$ and $\mathcal{F}_\phi(\mathcal{X}, \mu)$ the set of $\mu$-measurable functions $p$ defined over $\mathcal{X}$ and taking values in $I_\phi$. The $\phi$-entropy functional introduced in [4] is defined as

$$H_\phi(p) = -\int_{\mathcal{X}} \phi(p(x)) d\mu(x) \quad \forall p \in L^1_\phi = L^1_+ \cap \mathcal{F}_\phi$$

The second order differential of the entropy functional $H_\phi$ at $p$ in the direction of $f \in L^1_\phi$ is given by:

$$d^2 H_\phi(p; f, f) = -\int_{\mathcal{X}} \phi''(p(x))(f(x))^2 d\mu(x)$$

We now introduce the set $\theta = (\theta_1, ..., \theta_n)$ of $n$ real continuous parameters defining a manifold $\Theta$ embedded in $\mathbb{R}^n$ and consider the subset $\mathcal{F}_\Theta$ of $\mathcal{P}_\phi = \mathcal{P} \cap \mathcal{F}\phi$:

$$\mathcal{F}_\Theta = \{p(.|\theta) \in \mathcal{P}_\phi : \theta \in \Theta\}$$

It is a family of probability density functions of the random variable $x \in \mathcal{X}$. Moreover, it will be assumed in the following that $\forall i, j = 1, ..., n$:

1. $\frac{\partial p(x|\theta)}{\partial \theta_i}$ is defined $\forall x \in \mathcal{X}$ and $\forall \theta \in \Theta$
2. $\int_{\mathcal{X}} \frac{\partial p(x|\theta)}{\partial \theta_i} d\mu(x) = 0$
3. the matrix $\mathbb{E}_\theta \left[ \frac{\partial \log p(x|\theta)}{\partial \theta_i} \frac{\partial \log p(x|\theta)}{\partial \theta_j} \right]$ is positive definite $\forall \theta \in \Theta$

We wish to quantify the second variation of the entropy functional in the direction of a vector $dp(\theta)$ of the tangent space of $\Theta$, with

$$dp(\theta) = \sum_{i=1}^{n} \frac{\partial p(.|\theta)}{\partial \theta_i} d\theta_i$$

denoting the first order approximation of the difference between the densities associated with the parameter points $(\theta_1, ...\theta_n)$ and $(\theta_1 + d\theta_1, ...\theta_n + d\theta_n)$. Hence,

$$d^2\{H_\phi(p)\}(\theta) = -\int_{\mathcal{X}} \phi''(p(x))(dp(\theta))^2 d\mu(x)$$

Under the assumption that $\phi$ is convex in $I_\phi$

$$ds^2_\phi(\theta) = -d^2\{H_\phi(p)\}(\theta) = \sum_{i,j=1}^{n} g_{ij}^{(\phi)} d\theta_i d\theta_j$$

with

$$g_{ij}^{(\phi)} = \int_{\mathcal{X}} \phi''(p(x)) \frac{\partial p(x|\theta)}{\partial \theta_i} \frac{\partial p(x|\theta)}{\partial \theta_j} d\mu(x)$$

defines a positive definite form on the tangent space and thus gives a Riemannian metric on $\Theta$, known as the $\phi$-entropy metric.

*Proof.* This follows from the expansion of $(dp(\theta))^2$ and the linearity of the integral. $\square$

The $n \times n$ matrix $g_{ij}^{(\phi)}$ is the $\phi$-entropy matrix. The line element $ds = (ds^2_\phi(\theta))^{1/2}$ is easily seen to be invariant under transformation of $\theta$. Consequently, $g_{ij}^{(\phi)}(\theta)$ is a second order covariant symmetric tensor.

Various possible choices for the entropy function $\phi$ have been proposed. We concentrate, in this report, on the Shannon entropy associated with $\phi(p) = p \log p \quad \forall p \in \mathcal{F}_\Theta$. Then,

$$H(p) = -\int_{\mathcal{X}} p(x|\theta) \log p(x|\theta) d\mu(x)$$

and the components of the information matrix, known as the Fisher information matrix, become

$$g_{ij}(\theta) = \int_{\mathcal{X}} \frac{\partial \log p(x|\theta)}{\partial \theta_i} \frac{\partial \log p(x|\theta)}{\partial \theta_j} p(x|\theta) d\mu(x) = \mathbb{E}_\theta \left[ \frac{\partial \log p(x|\theta)}{\partial \theta_i} \frac{\partial \log p(x|\theta)}{\partial \theta_j} \right]$$

## 1.2 Information Geodesic Distance

The geodesic distance $\mathcal{D}$ induced by the information metric $ds^2$ previously derived from the Fisher information matrix was investigated for certain parametric distributions in [2], [4], [19], [3], [8] and others. More recently, Calvo and Oller [5] derived an explicit solution of the information geodesic equation for the general multivariate normal model.

We recall that, if $t \mapsto \theta(t)$, $t_1 \leq t \leq t_2$ denotes a curve segment in $\Theta$ between two parametrized distributions $\theta^1, \theta^2$, its length is expressed as

$$\mathcal{D}(\theta^1, \theta^2) = \left| \int_{t1}^{t2} \left( \sum_{i,j=1}^{n} g_{ij}(\theta) \frac{d\theta_i}{dt} \frac{d\theta_j}{dt} \right)^{1/2} dt \right| \tag{1}$$

A geodesic curve is characterized by the fact that it minimizes the information geodesic distance $\mathcal{D}(\theta^1, \theta^2)$, e.g. it is solution of the Euler-Lagrange equations:

$$\sum_{i=1}^{n} g_{ik}(\theta) \frac{d^2\theta_i}{dt^2} + \sum_{i,j=1}^{n} \Gamma_{ijk} \frac{d\theta_i}{dt} \frac{d\theta_j}{dt} = 0 \quad \forall k = 1, ..., n \tag{2}$$

where $\Gamma_{ijk}$ are the Christoffel symbols of the first kind. Solving the Euler-Lagrange equations and the evaluation of the geodesic distance $\mathcal{D}$ constitute, in general, a difficult task.

## 1.3 Application to the Multivariate Normal Model

Our ultimate goal being to define statistics on multivariate normal distributions and to apply it to diffusion tensor images, we now concentrate on the space $S^+(m, \mathbb{R})$, endowed with the information metric $ds^2$. $S^+(m, \mathbb{R})$ denotes the set of $m \times m$ real symmetric positive-definite matrices. Its elements are used to describe the covariance matrices of the fixed-mean (denoted by $\xi \in \mathbb{R}^m$) normal model. Through the mapping that associates to each $\Sigma \in S^+(m, \mathbb{R})$ its components $\sigma_{ij}$, $i \leq j$, $i, j = 1, ..., m$, we see that $S^+(m, \mathbb{R})$ is isomorphic to $\mathbb{R}^n$, $n = \frac{1}{2}m(m+1)$. Hence, the parameter space $\Theta$, whose elements are $\theta = (\sigma_{ij})$, $i \leq j$, is embedded in $\mathbb{R}^{\frac{1}{2}m(m+1)}$.

We denote by $E_{ij}$, $i \leq j, i, j = 1, ..., m$ the canonical basis of the tangent space $TS^+$ (e.g. the space of vector fields). We equally denote by $E^*_{ij}$, $i \leq j, i, j = 1, ..., m$ the dual basis of the cotangent space $T^*S^+$ (e.g. the space of differential forms). These basis are given by:

$$E_{ij} = \begin{cases} 1_{ii} & , i = j \\ (1_{ij} + 1_{ji}) & , i \neq j \end{cases} \qquad E^*_{ij} = \begin{cases} 1_{ii} & , i = j \\ \frac{1}{2}(1_{ij} + 1_{ji}) & , i \neq j \end{cases}$$

For clarity of notations, we will drop the parameters $m$ and $\mathbb{R}$ in $S^+(m, \mathbb{R})$ when no ambiguity is possible.

Following [24], [3], [8], [5] and [11], we will characterize $S^+$ as an affine symmetric space for which closed form expressions are available for the solution of the geodesic equation 2 as well as for the information geodesic distance (also known as Rao's distance). Since, in the next sections, $S^+$ will be characterized as a homogeneous space, we first recall basic facts on Lie groups and Lie algebras (we refer the reader to [13] for more details).

**Notions on Lie Groups and Homogeneous Spaces**

- If $H$ is a subgroup of a group $G$, then for any $g \in G$, the left/right coset of $H$ are respectively defined as the subsets of $G$ of the form $\{gh : h \in H\}$ and $\{hg : h \in H\}$.
- If $G$ is a topological space, the set of left cosets of $H$, denoted $G/H$ is a homogeneous space.
- A homogeneous space is characterized by the fact that its has a group homomorphism $\tau(G)$ defining a transitive group action of a Lie group $G$ onto $G/H$ and assigning at each element $g$ of the group $G$ a mapping $\tau(g) : G/H \to G/H$ with action $aH \mapsto gaH \; \forall a \in G$.
- The tangent space of a Lie group $G$ at the identity has the structure of a Lie algebra $\mathfrak{g}$, e.g. a vector space with the bilinear mapping known as the Lie bracket $[.,.]$.

In our case $S^+$ can be expressed as the homogeneous space $G/H$ where the Lie subgroup $H$ is associated with the Lie subalgebra $\mathfrak{h} \subseteq \mathfrak{g}$. $G$ is the general linear group of real $m \times m$ matrices $GL(m, \mathbb{R})$ and $H$ is the special orthogonal group $SO(m)$.

The corresponding algebras are respectively $\mathfrak{gl}(m, \mathbb{R})$, the set of $m \times m$ real matrices, and $\mathfrak{so}(m)$, the set of $m \times m$ skew-symmetric matrices. Then, $GL(m, \mathbb{R})$ acts on $S^+(m, \mathbb{R})$ so that for each $g \in G, \Sigma \in S^+$, $\tau(g) : \Sigma \mapsto g\Sigma g^T$.

One important property of the information metric $ds^2$ for the general multivariate normal distributions $p(.|\theta) = \mathcal{N}(\xi, \Sigma)$, as stated in [5], is its invariance under the transformation on $S^+$ defined as:

$$(\xi, \Sigma) \mapsto (\xi, P\Sigma P^T)$$

where $P \in GL(m, \mathbb{R})$. We recall that $\xi \in \mathbb{R}^m$ is a fixed mean vector and $\Sigma$ the covariance matrix of the distribution.

The Rao's distance also exhibits this invariance property. Hence the invariance under transformations of $GL(m, \mathbb{R})$. The consequences of this nice property are twofold:
First, the usual Levi-Civita connection is torsion free. But second, and most important, $S^+$ being symmetric, there is a strong relationship between the exponential map Exp from Riemannian geometry, the exponential map exp $T_U S^+ \rightarrow S^+$ on Lie groups and the matrix exponential through:

$$\mathrm{Exp}_{\mathbb{U}}(X) = \exp(2X) = e^{(2X)}$$

where $U$ is the unit $m \times m$ matrix and the tangent space of $S^+$ is identified with $S(m, \mathbb{R})$, the space of $m \times m$ real symmetric matrices.

**Proposition 1.1.** *The Riemannian metric for the space of fixed-mean multivariate normal distributions is given by the twice covariant tensors*

$$g(E_{i_1,i_2}, E_{j_1,j_2}) = \frac{1}{2} tr(\Sigma^{-1} E_{i_1,i_2} \Sigma^{-1} E_{j_1,j_2}) \quad i_1, i_2, j_1, j_2 = 1, ..., m \tag{3}$$

*whose components are explicitly given in section 5.*

*Proof.* This follows from the heuristic definition of the metric tensor, proposed by Rao, which amounts to characterize the likelihood ratio of two infinitesimally close distributions of $S^+$, associated respectively to $\theta$ and $\theta + d\theta$. This ratio express the influence of replacing $\theta$ by $\theta + d\theta$ and is defined as

$$\frac{dp(\theta)}{p(\theta)} = \sum_{i=1}^{n} \frac{\partial p(x|\theta)}{\partial \theta_i} p^{-1}(x|\theta) d\theta_i = \sum_{i=1}^{n} \frac{\partial \log p(x|\theta)}{\partial \theta_i} d\theta_i$$

Its variance was proposed as a distance measure between two contiguous distributions, e.g. the metric tensor. As exposed in Appendix 1 of [2], for the multivariate normal model we have

$$\frac{\partial \log p(x|\Upsilon)}{\partial \Upsilon} = a - \frac{xx^T}{2}$$

where $\Upsilon = \Sigma^{-1}$ and $a \in \mathbb{R}$. The variance of this distribution in the directions $A$ and $B \in TS^+$ is then:

$$\begin{aligned} &\mathrm{Var}\left(\tfrac{dp(\Sigma)}{p(\Sigma)}\right)(A, B) = \tfrac{1}{4}\mathrm{Var}(xx^T)(A, B) \\ &= \tfrac{1}{4}\mathrm{Cov}(\mathrm{tr}(Axx^T), \mathrm{tr}(Bxx^T)) = \tfrac{1}{4}\mathrm{Cov}(\langle x, Ax\rangle, \langle x, Bx\rangle) \\ &= \tfrac{1}{4}\mathrm{Cov}(x^T Ax, x^T Bx) = \tfrac{1}{2}\mathrm{tr}(\Sigma A \Sigma B) \end{aligned} \tag{4}$$

Since $A$ and $B$ are linear combinations of the basis vectors, we only need to calculate the metric tensor for the canonical basis field $E_{ij}$, which yields the result:

$$g(E_{i_i,i_2}, E_{j_1,j_2}) = \frac{1}{2}\mathrm{tr}(\Sigma^{-1} E_{i_i,i_2} \Sigma^{-1} E_{j_i,j_2}) \quad i_1, i_2, j_1, j_2 = 1, ..., m$$

□

An explicit expression of the geodesic curves $\Sigma(t)$, $t \in [a,b] \subset \mathbb{R}$ determined by the following system [3] was derived Calvo and Oller in [5] (where the proof of the following can be found).

$$\left\{ \begin{array}{ll} (\Sigma^{-1}\dot{\Sigma}) + cc^T\Sigma & = 0 \\ c^T\Sigma c + \frac{1}{2}\mathrm{tr}((\Sigma^{-1}\dot{\Sigma})^2) & = 1 \end{array} \right. \quad c \in \mathbb{R}^n$$

As exposed in [5] and [18], the geodesic starting from $\Sigma(a) \in S^+$ in the direction

$$\dot{\Sigma}(a) = \Sigma(a)^{1/2} X \Sigma(a)^{1/2} \in S = TS^+$$

writes:

$$\Sigma(t) = \Sigma(a)^{1/2}\mathrm{exp}(tX)\Sigma(a)^{1/2} \in S^+ \tag{5}$$

which can also be obtained by using the group action previously denoted by $\tau(GL(m,\mathbb{R}))$ to recast the computation of any geodesic into the special case $\Sigma(a) = \mathbb{I}_m$ as it was proposed in [10].

The information geodesic distance $\mathcal{D}$ between any two element $\Sigma_1$ and $\Sigma_2$ of $S^+$ is given by the theorem (the proof is briefly sketched for the sake of clarity and is available in [2]):

**Theorem 1.2. (S.T. Jensen, 1976)**
*Consider the family of multivariate normal distributions with common mean vector $\xi$ but differing variance-covariance matrices $\Sigma$. The geodesic distance between two members of the family with variance-covariance matrices $\Sigma_1$ and $\Sigma_2$ is given by*

$$\mathcal{D}(\Sigma_1, \Sigma_2) = \frac{1}{2}\mathrm{tr}(\log^2(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2})) = \frac{1}{2}\sum_{i=1}^{m}\log^2(\lambda_i)$$

*where the $\lambda_i$ denote the $m$ eigenvalues of the determinantal equation $|\lambda\Sigma_2 - \Sigma_1| = 0$.*

*Proof.* Let us consider the singular value decomposition of $\Sigma = UDU^T$, then

$$d\Sigma = dUDU^T + UdDU^T + UDdU^T$$

and $g(d\Sigma, d\Sigma)$ can be written after some simplifications as:

$$a + \frac{1}{2}\mathrm{tr}(D^{-1}dDD^{-1}dD) = a + \frac{1}{2}\sum_{k=1}^{m}\left(\frac{d\lambda_k}{\lambda_k}\right)^2$$

where $a \in \mathbb{R} \geq 0$ and the $\lambda_k$ are the components of the diagonal matrix $D$. By using the invariance of the information matrix under one-to-one transformations of the distributions, it is possible to set $\Sigma_1 = D_1$ and $\Sigma_2 = \mathbb{I}$ and among all the curves $(U(t), D(t))$ between $(\mathbb{I},\mathbb{I})$ and $(\mathbb{I}, D_1)$, the shortest path is achieved by $(\mathbb{I}, D(t))$ to yield the distance $\frac{1}{2}\sum_{i=1}^{m}\log^2(\lambda_i)$.

□

**Properties of the Geodesic Distance:**
The distance $\mathcal{D}$ on $S^+$ defined above exhibits some nice properties that we hereafter summarize:

1. Positivity: $\mathcal{D}(\Sigma_1, \Sigma_2) \geq 0$, $\mathcal{D}(\Sigma_1, \Sigma_2) = 0 \Leftrightarrow \Sigma_1 = \Sigma_2$
2. Symmetry: $\mathcal{D}(\Sigma_1, \Sigma_2) = \mathcal{D}(\Sigma_2, \Sigma_1)$
3. Triangle inequality: $\mathcal{D}(\Sigma_1, \Sigma_3) \leq \mathcal{D}(\Sigma_1, \Sigma_2) + \mathcal{D}(\Sigma_2, \Sigma_3)$
4. Invariance under congruence transformations: $\mathcal{D}(\Sigma_1, \Sigma_2) = \mathcal{D}(P\Sigma_1 P^T, P\Sigma_2 P^T)$ $\forall P \in GL(m, \mathbb{R})$
5. Invariance under inversion: $\mathcal{D}(\Sigma_1, \Sigma_2) = \mathcal{D}(\Sigma_1^{-1}, \Sigma_2^{-1})$

The interested reader can find more details about this in the technical report by Förstner and Moonen [11].

## 1.4 Structure of the Space of Multivariate Normal Distributions

In order to further study the specificities of the space of multivariate normal distributions, we now define three important quantities from Riemannian geometry: the Riemann-Christoffel curvature tensor, the Ricci tensor and the Ricci scalar. But before being able to define these elements, we have to *choose* an affine connection which allows us to map any tangent space $T_\Sigma$ at $\Sigma$ to the tangent space $T_{\Sigma'}$ at $\Sigma'$. This affine mapping depends on the curve $\Sigma(t)$ connecting $\Sigma$ and $\Sigma'$. This way, even if $\Sigma(t)$ is a loop, the tangent space $T_\Sigma$ is mapped onto itself along this curve but, the origin of $T_\Sigma$ can be mapped onto another point (e.g. not itself). The change in the direction of a vector under this mapping characterizes the curvature of the manifold.

The canonical connection on a Riemannian manifold is the Levi-Civita connection (or covariance derivative). An important properties of this connection is that it is compatible with the metric. In other words, the covariant derivative of the metric is zero.

Finally the Levi-Civita connection coefficients are given by $n^3$ functions called the Christoffel symbols of the second kind $\Gamma_{ij}^k$ defined as:

$$\Gamma_{ij}^k = g^{km}\Gamma_{ijm} = \frac{1}{2} g^{km} \left( \frac{\partial g_{jm}}{\partial \theta_i} + \frac{\partial g_{im}}{\partial \theta_j} - \frac{\partial g_{ij}}{\partial \theta_k} \right)$$

But as we said, we need to choose the affine connection to be considered since this will influence the curvature properties of the manifold.
Amari [1] has introduced a one-parameter family of affine connections, known as the $\alpha$-connections, in order to better represent the intrinsic properties of the family of probability distributions. The $\alpha$-connections can be written:

$$\Gamma_{ijk}^{(\alpha)} = \Gamma_{ijk}^{(1)} + \frac{1-\alpha}{2} T_{ijk}$$

where the third-order symmetric tensor $T_{ijk}$ is defined as

$$T_{ijk} = \mathbb{E}_\theta \left[ \frac{\partial \log p(x|\theta)}{\partial \theta_i} \frac{\partial \log p(x|\theta)}{\partial \theta_j} \frac{\partial \log p(x|\theta)}{\partial \theta_k} \right]$$

We see that the 0-connection boils down to the Levi-Civita connection. We recall that it is the only one to be compatible with the metric, that is the only one by which the parallel transport of a vector does not affect its length. The $\alpha$-connections are not compatible with the metric for $\alpha \neq 0$.

Moreover, it can be shown ([3]) that, for any exponential family (a special type of distributions in which the multivariate normal model can be recast), the $\alpha$-Riemann-Christoffel curvature writes

$$R^{(\alpha)}_{ijkl} = (1 - \alpha^2) R_{ijkl}$$

thus giving a zero $\alpha$-curvature to the multivariate normal model for all $\alpha \neq 0$.
Because of their non-compatibility with the metric and their induced 0-curvature, the $\pm 1$-connection (respectively known as Efron and David) do not seem to be a good candidate for the following derivations of statistics on the space of multivariate normal distributions. We will indeed need our space to exhibit a non-positive sectional curvature in order to ensure the existence and uniqueness of the Riemannian barycenter. For this reason, we stick with the classical Levi-Civita connection in the remaining developments. We have to notice, however, that $\alpha$-connections with $\alpha \neq \pm 1$ will have to be investigated.

As shown in [24] and proved in [23], the Riemann-Christoffel curvature tensor corresponding to the information metric, for zero-mean multivariate normal distributions, with $X, Y, Z, V$ being elements of the canonical basis of vector fields and $\Sigma \in S^+$ is:

$$\begin{aligned} R_{ijkl} = R(X, Y, Z, V) = \ & \tfrac{1}{4} \quad \mathrm{tr}(Y\Sigma^{-1}X\Sigma^{-1}Z\Sigma^{-1}V\Sigma^{-1}) - \\ & \tfrac{1}{4} \quad \mathrm{tr}(X\Sigma^{-1}Y\Sigma^{-1}Z\Sigma^{-1}V\Sigma^{-1}) \end{aligned}$$

It has 1296 components but only 105 of which are independent (since we are in 6 dimensions).

**Remark:** We have restricted the expression of the Riemann-Christoffel tensor for the case of zero-mean multivariate normal distributions, e.g. for $S^+$. The previous developments were valid for the fixed-mean case. Expressions for normal distributions with general mean vector are available in [24].

The important point is that we can now compute the sectional curvature $\kappa$ of the manifold $S^+$ and verify that is actually non-constant and non-positive. A sectional curvature is associated to any two-dimensional subset of vectors $\mathcal{E}$ in the tangent space at $\Sigma$, $T_\Sigma$. It is defined as the Gaussian curvature of that hypersurface made of the set of points reached by the geodesics starting at $\Sigma$ in all the directions described by the plane $\mathcal{E}$.

Skovgaard has shown in [24] that, if we denote by $\rho_{ij}^2 = \frac{\sigma_{ij}^2}{\sigma_{ii}\sigma_{jj}}$ the correlation coefficient between the components of the covariance matrices $\Sigma = (\sigma_{ij})$, $i, j = 1, ..., m$, the sectional curvature at $\Sigma$ is given, for $i \neq j$, by

1. $\kappa(\mathcal{E}, \Sigma) = -\frac{\rho_{ij}^2}{1+\rho_{ij}^2} \quad \text{if}\, \mathcal{E} = \text{span}(E_{ii}, E_{jj})$
2. $\kappa(\mathcal{E}, \Sigma) = -\frac{1}{2} \quad \text{if}\, \mathcal{E} = \text{span}(E_{ii}, E_{ij})$

which indeed depends on $\Sigma$ and is non-positive.

We can derive the Ricci tensor, which is defined as a the contraction of the Riemann-Christoffel tensor, and which can be thought of as the object which measures the growth rate of the Riemannian metric. It is a symmetric, $m \times m$ covariant tensor. We recall that the Riemann-Christoffel tensor is expressed in term of the partial derivatives of the metric:

$$R_{ijk}^l = \frac{\partial \Gamma_{ik}^l}{\partial \theta_j} - \frac{\partial \Gamma_{ij}^l}{\partial \theta_k} + \Gamma_{ik}^m \Gamma_{mj}^l - \Gamma_{ij}^m \Gamma_{mk}^l$$

and the expression of the Ricci tensor follows:

$$Ricci_{ij} = R_{ijk}^k = R_{ijkl} g^{kl}$$

where $g^{kl}$ denotes the inverse of the metric. Summarizing everything, we have the expression for $R_{ijkl}$ and $g_{ij}$. Symbolic calculations easily lead to the components of the Ricci tensor in terms of the components of $\Sigma = (\sigma_{ij})$ A quite interesting point is that, by comparison of the Ricci tensor with the metric tensor, we can also deduce that the space of zero-mean multivariate normal distributions is, in general, not an Einstein space. It is indeed a space of non-constant curvature and hence does not verify the relation

$$Ricci_{ij} = \frac{\nu}{n} g_{ij}$$

where $\nu$ is the scalar curvature and $n$ the dimension of the space (e.g. $n = 6$).
Additionally, we can express the Ricci scalar curvature $\nu$, as the full trace of the Ricci tensor. It is also defined as twice the sum of all the sectional curvatures along all the 2-planes $\mathcal{E}$ and recalling that

$$\Sigma = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_2 & \theta_4 & \theta_5 \\ \theta_3 & \theta_5 & \theta_6 \end{bmatrix}$$

we have:

$$
\begin{aligned}
&\nu = g^{ij} Ricci_{ij} = \\
-0.03125( \quad & 51\, \theta_1^3\theta_4^3\theta_6^3 - 99\,(\theta_2^6\theta_6^3 + \theta_3^6\theta_4^3 + \theta_1^3\theta_5^6) - \\
& 161\,(\theta_1^2\theta_3^2\theta_6^2\theta_4^3 + \theta_1^3\theta_5^2\theta_4^2\theta_6^2 + \theta_1^2\theta_2^2\theta_4^2\theta_6^3) + \\
& 209\,(\theta_1^3\theta_5^4\theta_4\theta_6 + \theta_1\theta_2^4\theta_6^3\theta_4 + \theta_1\theta_3^4\theta_4^3\theta_6) - \\
& 221\,(\theta_1^2\theta_2^2\theta_5^4\theta_6 + 221\theta_1\theta_2^4\theta_6^2\theta_5^2 + \theta_1^2\theta_3^2\theta_4\theta_5^4 + \theta_1\theta_3^4\theta_4^2\theta_5^2 + \theta_2^2\theta_3^4\theta_4^2\theta_6 + \theta_2^4\theta_3^2\theta_6^2\theta_4) + \\
& 306\,\theta_1^2\theta_2\theta_3\theta_5\theta_4^2\theta_6^2 + 366\,(\theta_1^2\theta_2^2\theta_5^2\theta_4\theta_6^2 + \theta_1\theta_2^2\theta_3^2\theta_6^2\theta_4^2 + \theta_1^2\theta_3^2\theta_4^2\theta_6\theta_5^2) + \\
& 426\,\theta_1\theta_2^2\theta_5^2\theta_4\theta_6\theta_3^2 + 522\,(\theta_1^2\theta_2\theta_3\theta_5^5 + \theta_2^5\theta_3\theta_6^2\theta_5 + \theta_2\theta_3^5\theta_5\theta_4^2) + \\
& 768\,\theta_2^3\theta_3^3\theta_5^3 - 796\,(\theta_1\theta_2\theta_3^3\theta_5\theta_6\theta_4^2 + \theta_1\theta_2^3\theta_3\theta_6^2\theta_4\theta_5 + \theta_1^2\theta_2\theta_3\theta_5^3\theta_4\theta_6) + \\
& 940\,(\theta_2^3\theta_6\theta_5\theta_3^3\theta_4 + \theta_1\theta_2^3\theta_3\theta_6\theta_5^3 + \theta_1\theta_2\theta_3^3\theta_5^3\theta_4) - \\
& 1056\,(\theta_1\theta_2^2\theta_5^4\theta_3^2 + \theta_2^4\theta_6\theta_5^2\theta_3^2 + \theta_2^2\theta_3^4\theta_5^2\theta_4))/\Delta^3
\end{aligned}
$$

where $\Delta = |\mathbf{\Sigma}| = \theta_1\theta_4\theta_6 - \theta_1\theta_5^2 - \theta_2^2\theta_6 + 2\theta_2\theta_3\theta_5 - \theta_3^2\theta_4$.

# 2 Statistics on Multivariate Normal Distributions

## 2.1 Intrinsic Mean

Following a recent work by Moakher [18], we propose a new gradient descent algorithm for the computation of the intrinsic mean distribution of a set of zero-mean multivariate normal distributions. It relies on the classical definition of the Riemannian center of mass and makes use of the geodesic equations to derive a manifold constrained numerical integrator and thus ensure that each step forward of the gradient descent stays within the space $S^+$. As we will show in the numerical experiments, this method is very efficient and usually converges in just a few iterations.

We, of course, seek to estimate the empirical mean as proposed by Fréchet [12], Karcher [14], Pennec [20] or Moakher [17] :

**Definition 2.1.** *The Normal distribution parametrized by $\hat{\Sigma} \in S^+(m, \mathbb{R})$ and defined as the empirical mean of $N$ distributions $\Sigma_k$, $k = 1, ..., N$, achieves a local minimum of the function $\mu : S^+(m, \mathbb{R}) \to \mathbb{R}^+$ known as the empirical variance and defined as*

$$\mu(\Sigma_1, ..., \Sigma_N) = \frac{1}{N}\sum_{k=1}^{N} \mathcal{D}^2(\Sigma, \Sigma_k) = \mathbb{E}[\mathcal{D}^2(\Sigma, \Sigma_k)]$$

Karcher proved in [14] that such a mean, known as the Riemannian barycenter, exists and is unique for manifold of non-positive sectional curvature. This was shown to be the case for $S^+$.

In order to derive our gradient descent algorithm, we rely on the well-known following remarks (see [7] and references therein for more details): We want to derive a flow evolving an initial guess $\hat{\Sigma}(0)$ toward the mean of a set of elements of $S^+$. If we denote by $\hat{\Sigma}(s), s \in [0, \infty)$ the family of solution of

$$\partial_s \hat{\Sigma}(s) = V(\hat{\Sigma}(s))$$

where $V$ is the direction of evolution, we have

$$\begin{aligned} & \hat{\Sigma}(s) \in S^+(m, \mathbb{R}),\ \forall s > 0 \\ \Leftrightarrow\ & \hat{\Sigma}(0) \in S^+(m, \mathbb{R}) \text{ and } V(\hat{\Sigma}(s)) \in T_{\hat{\Sigma}(s)} S^+(m, \mathbb{R}) = S(m, \mathbb{R}),\ \forall s > 0 \end{aligned}$$

We thus identify $V$ with the opposite of the covariant derivative of our objective function $\mu(\Sigma_1, ..., \Sigma_N)$ which is shown to be (see [18]):

$$\nabla \mu = \frac{\hat{\Sigma}(s)}{N} \sum_{k=1}^{N} \log(\Sigma_k^{-1} \hat{\Sigma}(s)) \tag{6}$$

Hence the evolution

$$\partial_s \hat{\Sigma}(s) = -\frac{\hat{\Sigma}(s)}{N} \sum_{k=1}^{N} \log(\Sigma_k^{-1} \hat{\Sigma}(s)) \tag{7}$$

Numerical Implementation:
But, as mentioned in [7], the corresponding numerical implementation has to be dealt with carefully and we have to build a step-forward operator $\mathbf{K}_{dt}$ such that the discrete flow

$$\Sigma_{l+1} = \mathbf{K}_{dt}(\Sigma_l),\ \Sigma_0 \in S^+$$

provides an intrinsic (or consistent as qualified in [7]) approximation of evolution equation 7 ($l$ denotes the discrete version of the continuous family parameter $s$). As we have previously pointed out, a closed form for the geodesics of $S^+$ does exist. Hence we can directly integrate along the geodesics of the space without having to use the exponential map to re-project, after each integration step, the element obtained by moving in the direction given by $-\nabla\mu$ and we have the following proposition:

**Proposition 2.2.** *For any $\Sigma(t) \in S^+(m, \mathbb{R})$, $t \in [a, b]$ and any tangent vector $V = -\nabla\mu \in S(m, \mathbb{R})$, an intrinsic (or consistent) approximation of the flow 7 can be obtained by using the step-forward operator*

$$\mathbf{K}_{dt}(\Sigma_l) = \Sigma_l^{1/2} \exp\left(-dt \Sigma_l^{-1/2} \nabla\mu \Sigma_l^{-1/2}\right) \Sigma_l^{1/2} \tag{8}$$

*Proof.* The geodesic starting from $\Sigma(t)$ and pointing in direction $V = \Sigma(t)^{1/2} X \Sigma(t)^{1/2}$, $X \in M(m, \mathbb{R})$ is given by

$$\Sigma(t + dt) = \Sigma(t)^{1/2} \exp\left([t + dt]X\right) \Sigma(t)^{1/2} \quad \forall dt \in [0, 1]$$

it readily follows that

$$\begin{aligned}\Sigma(t+dt) &= \Sigma(t)^{1/2}\exp(tX)\Sigma(t)^{1/2}\Sigma(t)^{-1/2}\exp(dtX)\Sigma(t)^{1/2}\\ &= \Sigma(t)\Sigma(t)^{-1/2}\exp(dtX)\Sigma(t)^{1/2}\end{aligned}$$

If $V$ is identified with the opposite of the covariant derivative of any objective functional $\mu$ (which is in $S(m,\mathbb{R})$), we obtain:

$$X = -\Sigma(t)^{-1/2}\nabla\mu\Sigma(t)^{-1/2}$$

and

$$\Sigma(t+dt) = \Sigma(t)\Sigma(t)^{-1/2}\exp(-dt\Sigma(t)^{-1/2}\nabla\mu\Sigma(t)^{-1/2})\Sigma(t)^{1/2} \tag{9}$$

identifying $\Sigma(t)$ with the current estimate of the mean $\Sigma_l$ yields the result.

□

QUESTION: How does this numerical scheme compare to the gradient descent associated to the following operator?

$$\mathbf{K}_{dt}(\Sigma_l) = \Sigma_l - dt\nabla\mu \tag{10}$$

**Proposition 2.3.** *The extrinsic step-forward operator 10 is a first order approximation of the operator of Proposition 2.2.*

*Proof.* Recalling that the matrix exponential writes as the following power series:

$$\exp X = \sum_{n=0}^{\infty}\frac{X^n}{n!} = \mathbb{I} + X + \frac{XX}{2} + \frac{XXX}{6} + ...$$

a first order expansion of 8 yields:

$$\begin{aligned}\mathbf{K}_{dt}(\Sigma_l) &= \Sigma_l\Sigma_l^{-1/2}(\mathbb{I} - dt\Sigma_l^{-1/2}\nabla\mu\Sigma_l^{-1/2})\Sigma_l^{1/2}\\ &= \Sigma_l - dt\Sigma_l\Sigma_l^{-1}\nabla\mu\\ &= \Sigma_l - dt\nabla\mu\end{aligned}$$

□

The exponential map is defined from $[0,1]$ onto $S^+$. The optimal time step $dt$ is then equal to 1 and the optimal step-forward operator is expressed as follows

$$\mathbf{K}(\Sigma_l) = \Sigma_l^{1/2}\exp(-\Sigma_l^{-1/2}\nabla\mu\Sigma_l^{-1/2})\Sigma_l^{1/2}$$

This will be illustrated in the section dedicated to numerical experiments.

CONCLUSION:
We now come back to the derivation of a numerical algorithm to estimate the Riemannian

barycenter $\hat{\Sigma}$ of a set of parametrized multivariate normal distributions. We simply make use of the explicit expression of the covariant derivative $\nabla\mu$ given in 6. This yields the step-forward operator:

$$\mathbf{K}_{dt}(\hat{\Sigma}_l) = \hat{\Sigma}_l^{1/2} \exp\left(-dt\, \frac{\hat{\Sigma}_l^{1/2}}{N} \sum_{k=0}^{N} \log(\Sigma_k^{-1}\hat{\Sigma}_l)\hat{\Sigma}_l^{-1/2}\right)\hat{\Sigma}_l^{1/2} \tag{11}$$

whose associated flow converges toward the barycenter for any initial guess $\hat{\Sigma}_0$.

## 2.2 Intrinsic Covariance Matrix and Principal Modes

We propose a new algorithm for the computation of the intrinsic empirical covariance matrix of a set of $N$ zero-mean multivariate normal distributions. We follow [6] where this problem was addressed in the infinite dimensional case of a set of planar closed curves, and [10] where the Logarithmic mapping had to be used. Our objective will be to derive an intrinsic numerical scheme for the estimation of the covariance matrix $\mathbf{\Lambda}$ relative to the empirical mean $\hat{\Sigma}$ of $N$ normal distributions and using the explicit solution of the geodesic equation. As we consider the 6-dimensional manifold of parametrized normal laws, we will of course end up with $\mathbf{\Lambda} \in S^+(6,\mathbb{R})$ acting on the space $S(6,\mathbb{R})$.
As defined in [6] and [20], we associate to each of $N$ normal distribution $\Sigma_k$ the unique tangent vector $\beta_k \in S(m,\mathbb{R})$ (seen as an element of $\mathbb{R}^m$) such that the empirical mean $\hat{\Sigma}$ is mapped onto $\Sigma_k$ by the exponential map $\mathrm{Exp}_{\hat{\Sigma}}(\beta_k)$ (see figure 1). We then have the following definition:

**Definition 2.4.** *Given $N$ elements of $S^+(m,\mathbb{R})$ and a mean value $\hat{\Sigma}$, the empirical covariance matrix relative to $\hat{\Sigma}$ is defined as:*

$$\Lambda_{\hat{\Sigma}} = \frac{1}{N}\sum_{k=1}^{N} \beta_k \beta_k^T$$

We identify the $\beta_k$ with the covariant derivative of the squared distance function $\nabla\mathcal{D}^2(\hat{\Sigma},\Sigma_k)$. As detailed in [18], we have $\nabla\mathcal{D}^2(\hat{\Sigma},\Sigma_k) = \hat{\Sigma}\log(\Sigma_k^{-1}\hat{\Sigma})$.

Once we have computed $\Lambda$, it is fairly easy, and instructive in order to further understand the structure of our subset of normal distributions, to compute the eigenvalues and eigenvectors of the covariance matrix. Fletcher et al. [9] generalized the notion of Principal Component Analysis to Lie groups by seeking geodesic submanifolds, by analogy with the lower-dimensional linear subspaces of PCA, that maximize the projected variance of the data. This actually amounts to characterize the tangent space at the mean element $\hat{\Sigma}$ through classical PCA in order to construct an orthogonal basis of tangent vectors $v_k$, $k = 1,...,d \le m$ that can be used to generate $l$-dimensional subspaces $V_l = \mathrm{span}(v_1,...,v_l)$, $l \le m$ that maximize the projected variance. The $v_k$ are defined as the set of eigenvectors of the covariance matrix $\Lambda$. However, since we have available the geodesic equations in $S^+$, we have a closed-form to

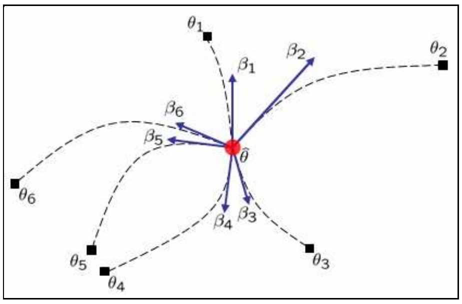


Figure 1: Depiction of the velocity field $\beta_k$ at the empirical mean $\hat{\Sigma}$

generate elements of any geodesic submanifold $H_k$ defined as the exponential mapping the $V_k$.

Indeed, we can define any linear combination of the $v_k$, $v = \sum_{k=1}^{d} \alpha_k v_k \in S(m, \mathbb{R})$ and then compute the unique element $C$ of $S^+(m, \mathbb{R})$ reached by following the geodesic emanating from $\hat{\Sigma}$ in the direction $v$ and computed as:

$$C = \hat{\Sigma}^{1/2} \exp(\hat{\Sigma}^{-1/2} \sum_{k=1}^{d} \alpha_k v_k \hat{\Sigma}^{-1/2}) \hat{\Sigma}^{1/2} \tag{12}$$

## 2.3 A Normal Distribution on Multivariate Normal Distributions

Our last contribution makes use of the various quantities derived up to that point and fully exploits the information they provide in order to derive the expression of a normal distribution on $S^+$. We thus proceed by plugging the appropriate quantities in the generalization of the normal distribution to Riemannian manifolds proposed in [20] for sufficiently concentrated probability density functions, e.g. for small covariance matrices.
We are motivated by a very important application in medical image analysis, namely the segmentation of diffusion Magnetic Resonance Images. The authors have already proposed several different approaches to tackle this issue [16], [22], [15]. However, since the essence of these approaches is to maximize the log-likelihood of a (normal) distribution over the objects of interest, e.g. Diffusion Tensors and since those tensors are naturally interpreted as the covariance matrices of zero-mean normal distributions, the following generalized law will be of great help and is currently being applied to that task.

Following Theorem 4 proved in [20], we have the following proposition :

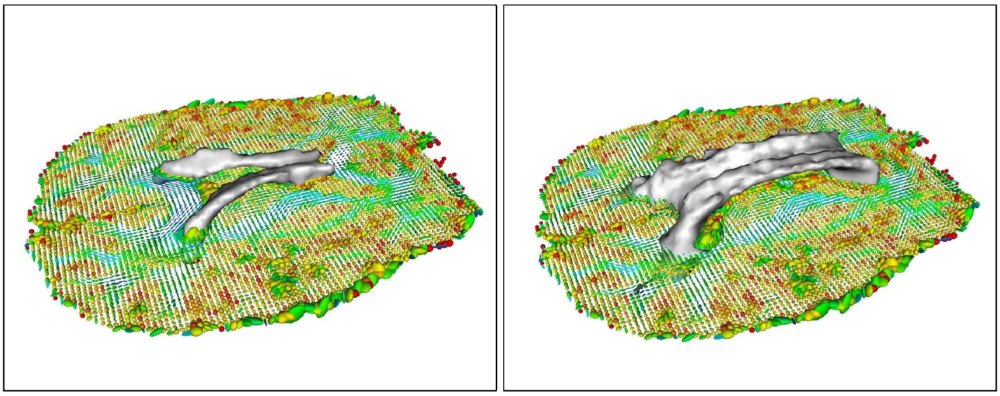

Figure 2: Examples of segmentation on diffusion tensor MRI by surface evolution ([LEFT] ventricles [RIGHT] *corpus callosum*)

**Proposition 2.5.** *The normal distribution in* $S^+(m, \mathbb{R})$ *for an* $m \times m$ *covariance matrix* $\Lambda$ *of small variance* $\sigma^2 = \mathrm{tr}(\Lambda)$ *is of the form*

$$k.\exp\frac{-\beta\gamma\beta^T}{2}$$

*where*

- $\beta$ *is defined as* $\nabla\mathcal{D}^2(\hat{\Sigma}, \Sigma)$, *as in section 2.2*
- *The normalization constant:* $k = \frac{1+O(\sigma^3)+\epsilon(\frac{\sigma}{r})}{\sqrt{(2\pi)^m|\Lambda|}}$, *with* $\Lambda$ *defined as in section 2.2*
- *The concentration matrix:* $\gamma = \Lambda^{-1} - \frac{Ricci}{3} + O(\sigma) + \epsilon(\frac{\sigma}{r})$, *with* $\Lambda$ *defined as in section 2.2 and* $Ricci$ *as in section 1.4*

$r$ *is the injection radius at* $\hat{\Sigma}$ *and* $\epsilon$ *is such that* $\lim_{o^+} x^{-p}\epsilon(x) = 0 \; \forall p \in \mathbb{R}^+$.

# 3 Numerical Experimentations

We propose, in Algorithms 1 and 2, the Matlab™ implementation of the algorithms for the intrinsic mean and covariance matrix. The computation of the modes of variation amounts to perform a Singular Value Decomposition of the covariance matrix and to use equation 12 to estimate the modes. The implementation is straightforward.
Regarding, the computation of the Ricci tensor, we do not provide the explicit expression of its components since it takes too much space but the use of Maple™ with the `tensor` and `codegeneration` packages is recommended and easily provides $C$ source code. Finally, the Ricci scalar curvature is obtained directly from the expression given at the end of section 1.4.

**Algorithm 1** Matlab Code for Intrinsic Mean

```
% Inputs
% S    : Array of 3x3 symmetric positive definite matrices (component-wise)
% M0   : Initial guess of the mean
% dt   : Time step
% niter: Number of iterations
%
% Output
% M    : Estimated Mean matrix
function [M] = intrinsicMean(S,M0,dt,niter)[N,w,h] = size(S);
[n,m] = size(M0);
M = M0;
for iter=1:niter
    V = zeros(n,m);
    for i=1:N
        V = V + logm(inv(reshape(S(i,:,:),3,3))*M);
    end;
    M = sqrtm(M)*expm(-dt*sqrtm(M)*(V./N)*inv(sqrtm(M)))*sqrtm(M);
end;
```

**Algorithm 2** Matlab Code for Intrinsic Covariance Matrix

```
% Inputs
% S    : Array of 3x3 symmetric positive definite matrices (component-wise)
% M    : Estimated mean
%
% Output
% C    : Estimated covariance matrix
function [C] = intrinsicCovMat(S,M)
[N,h,w] = size(S);
C = zeros(6,6);
W = zeros(6,1);
for i=1:N
    P = reshape(S(i,:,:),3,3);
    V = M*logm(inv(P)*M);
    W(1) = V(1,1); W(2) = V(1,2); W(3) = V(1,3);
    W(4) = V(2,2); W(5) = V(2,3); W(6) = V(3,3);
    C = C + W*W';
end;
C = C./N;
```

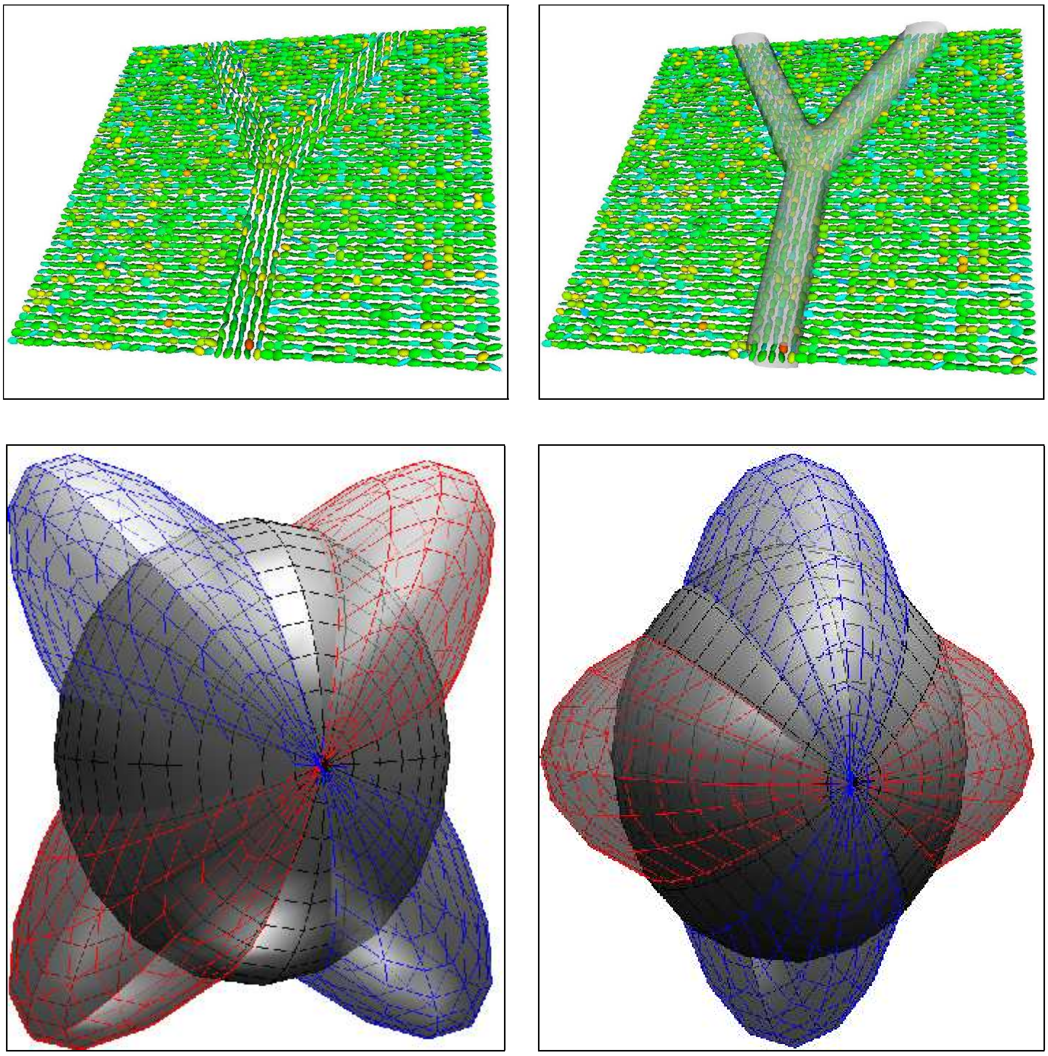

Figure 3: [TOP] Synthetic tensor field ([left] original noisy field, [right] segmentation of the shape of interest), [BOTTOM] First and second modes of variation superimposed on the mean tensor (black:mean, red:$\alpha_k = +2$ in equation 12 , blue:$\alpha_k = -2$ in equation 12)

| Elements | 10 | 100 | 500 | 1000 | 5000 | 10000 | 50000 |
|---|---|---|---|---|---|---|---|
| Time (in $s$) | 0.0665 | 0.4143 | 1.9685 | 3.8459 | 19.5005 | 39.4221 | 193.7162 |

Table 1: Time of convergence in terms of the number of elements (on a 2,4GHz PIV)

We now proceed to numerical experimentations on synthetic and real datasets.
We have first applied the definition of the geodesic distance and its associated variance to perform the segmentation of cerebral anatomical structures by fitting 1-dimensional normal models on the distributions of the geodesic distances. The interested reader is referred to [16] for more details. Figure 2 presents the results of the segmentation for the ventricles and the well-known structure of the white matter, the *corpus callosum*.

We then test the Algorithms 1 and 2 on a subset of the three-dimensional synthetic tensor field shown on figure 3. The data was generated so that some tensors follow a $Y$ pattern. The rest of the volume was filled with tensors aligned in the same direction. Random symmetric matrices, with entries following a normal law of zero mean and variance 1 were made positive-definite by application of the matrix exponential and voxelwise added to the volume. We actually extracted the tensors inside the $Y$ shape (about 2000) and computed the mean, covariance matrix and first two modes of variations. The result are presented on figure 3. For the sake of clarity, we only have represented the first two modes for $\alpha_k = \pm 2$ in equation 12. It turns out that the first two eigenvectors of the covariance matrix are dominant, with almost equal eigenvalues. The first mode seems to capture the orientational variation of the dataset whereas the second one seems to encapsulate the variation of size of the tensors.

Next, we have conducted some performance analysis of the gradient descent algorithm used to estimate the intrinsic mean. By running the algorithm repeatedly with different initial guesses, we have ensured that it was not affected by that parameter since it converged every time in no more than 4 iterations. This was highly reproducible and tested with a known mean and noise generated as previously. On that basis, we present on figure 1, a table showing the evolution of the time of convergence in terms of the number of elements for which the mean is to be estimated. As expected, this is a perfectly linear complexity.

Finally, we would like to show interesting results on the computation of the Ricci scalar on real Diffusion Tensor Magnetic Resonance Imaging. We recall that this particular MR modality produces, at each voxel, a tensor that is nothing but the covariance matrix of the normal law approximating the average motion of water molecules. On figure 4 are presented the fractional anisotropy and the scalar curvature computed on a $64 \times 64 \times 24$ DTI volume. The fractional anisotropy is a combination of the eigenvalues of the diffusion tensor which is well known to highlight the regions where a high density of neural fibers are present. However, this measure seems to be somehow sensitive to noise and highly unreliable at voxel where crossings of fibers occur. Thinking of our DTI data as a field of normal distributions,

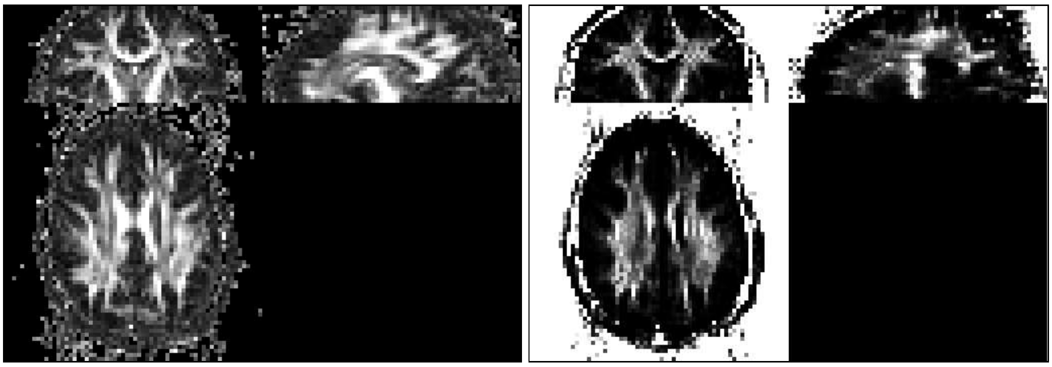

Figure 4: [LEFT] Fractional anisotropy, [RIGHT] Scalar curvature computed on real DTI data

it is easy to see, from the definition of the scalar curvature $\nu$, that this quantity will vary with the anisotropy of the covariance matrix of each normal distribution. Indeed, figure 4 [RIGHT] shows that the scalar curvature has a quite better discriminative power than the fractional anisotropy FA. Well-known structures of the white-matter such as the *corpus callosum* or the *corona radiata* are indeed clearly highlighted in figure 4 [RIGHT] (we recall that the scalar curvature is negative, so that a black voxel exhibits the most negative curvature). This is a promising new measure of anisotropy for DTI that we are currently evaluating.

# 4 Conclusion

We have presented a geometric approach to the statistical analysis of multivariate normal distributions. We have developed novel algorithms for the estimation of the mean, covariance matrix and the modes of variation. We have also introduced explicit expressions for the Riemann-Christoffel and Ricci tensors as well as for the Ricci scalar curvature for the space of zero-mean multivariate normal distributions. All these contributions have been used in order to derive and fully characterize a generalized normal law on the space of zero-mean multivariate normal distributions. We have shown promising results on synthetic and real datasets. One of our goal is now to exploit all the information provided by those statistics in order to better understand and analyze the data obtained by diffusion tensor imaging.

**Acknowledgments:** The authors would like to thank M. Moakher (National Engineering School at Tunis, Tunisia) for his fruitful comments on parts of the present research report. They also thank K. Ugurbil, S. Lehericy and G. Sapiro (Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, USA) for their valuable collaboration and expertise to acquire the data used in this report on figure 4 as well as J.F. Mangin and J.B. Poline (CEA SHFJ, Orsay, France) for providing us with the data used on figure 2.

# 5 Appendix 1: Components of the Riemannian metric

$g_{11} = \left(\theta_4\theta_6 - \theta_5^2\right)^2 / 2\Delta^2$

$g_{12} = \left(\theta_4\theta_6 - \theta_5^2\right)\left(\theta_2\theta_6 - \theta_3\theta_5\right) / \Delta^2$

$g_{13} = \left(\theta_4\theta_6 - \theta_5^2\right)\left(\theta_2\theta_5 - \theta_3\theta_4\right) / \Delta^2$

$g_{14} = \left(\theta_2\theta_6 - \theta_3\theta_5\right)^2 / 2\Delta^2$

$g_{15} = -\left(\theta_2\theta_6 - \theta_3\theta_5\right)\left(\theta_2\theta_5 - \theta_3\theta_4\right) / \Delta^2$

$g_{16} = -\left(\theta_2\theta_5 - \theta_3\theta_4\right)^2 / 2\Delta^2$

$g_{22} = \left(-\theta_6^2\theta_2^2 - 2\theta_3^2\theta_5^2 + 2\theta_6\theta_2\theta_3\theta_5 - \theta_1\theta_4\theta_6^2 + \theta_6\theta_1\theta_5^2 + \theta_6\theta_3^2\theta_4\right) / \Delta^2$

$g_{23} = \left(-\theta_2^2\theta_6\theta_5 + 2\theta_2\theta_6\theta_3\theta_4 - \theta_3^2\theta_5\theta_4 - \theta_5\theta_1\theta_4\theta_6 + \theta_1\theta_5^3\right) / \Delta^2$

$g_{24} = \left(-\theta_1\theta_6 + \theta_3^2\right)\left(\theta_2\theta_6 - \theta_3\theta_5\right) / \Delta^2$

$g_{25} = \left(-\theta_4\theta_3\theta_1\theta_6 + \theta_4\theta_3^3 - \theta_6\theta_2^2\theta_3 + 2\theta_6\theta_2\theta_1\theta_5 - \theta_5^2\theta_3\theta_1\right) / \Delta^2$

$g_{26} = \left(\theta_2\theta_5 - \theta_3\theta_4\right)\left(-\theta_1\theta_5 + \theta_3\theta_2\right) / \Delta^2$

$g_{33} = \left(\theta_4^2\theta_3^2 - 2\theta_4\theta_2\theta_3\theta_5 + 2\theta_2^2\theta_5^2 - \theta_1\theta_4\theta_5^2 - \theta_2^2\theta_6\theta_4 + \theta_6\theta_4^2\theta_1\right) / \Delta^2$

$g_{34} = \left(\theta_2\theta_6 - \theta_3\theta_5\right)\left(-\theta_1\theta_5 + \theta_3\theta_2\right) / \Delta^2$

$g_{35} = \left(\theta_4\theta_2\theta_3^2 - 2\theta_4\theta_1\theta_3\theta_5 + \theta_2\theta_1\theta_5^2 - \theta_2^3\theta_6 + \theta_4\theta_6\theta_2\theta_1\right) / \Delta^2$

$g_{36} = -\left(-\theta_1\theta_4 + \theta_2^2\right)\left(\theta_2\theta_5 - \theta_3\theta_4\right) / \Delta^2$

$g_{44} = -\left(-\theta_1\theta_6 + \theta_3^2\right)^2 / 2\Delta^2$

$g_{45} = -\left(-\theta_1\theta_6 + \theta_3^2\right)\left(-\theta_1\theta_5 + \theta_3\theta_2\right) / \Delta^2$

$g_{46} = -\left(-\theta_1\theta_5 + \theta_3\theta_2\right)^2 / 2\Delta^2$

$g_{55} = \left(\theta_4\theta_1\theta_3^2 + 2\theta_2\theta_1\theta_3\theta_5 - \theta_1^2\theta_5^2 - 2\theta_2^2\theta_3^2 + \theta_6\theta_2^2\theta_1 - \theta_4\theta_6\theta_1^2\right) / \Delta^2$

$g_{56} = -\left(-\theta_1\theta_4 + \theta_2^2\right)\left(-\theta_1\theta_5 + \theta_3\theta_2\right) / \Delta^2$

$g_{66} = \left(-\theta_1\theta_4 + \theta_2^2\right)^2 / 2\Delta^2$

where the determinant of the covariance matrix is:
$\Delta = |\mathbf{D}| = \theta_1\theta_4\theta_6 - \theta_1{\theta_5}^2 - \theta_2^2\theta_6 + 2\theta_2\theta_3\theta_5 - \theta_3^2\theta_4$

## References

[1] S.I. Amari. *Differential-Geometrical Methods in Statistics.* Lectures Notes in Statistics. Springer-Verlag, 1990.

[2] C. Atkinson and A.F.S. Mitchell. Rao's distance measure. *Sankhya: The Indian Journal of Stats.*, 43(A):345–365, 1981.

[3] J. Burbea. Informative geometry of probability spaces. *Expositiones Mathematica*, 4:347–378, 1986.

[4] J. Burbea and C.R. Rao. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *Journal of Multivariate Analysis*, 12:575–596, 1982.

[5] M. Calvo and J.M. Oller. An explicit solution of information geodesic equations for the multivariate normal model. *Statistics and Decisions*, 9, 1991.

[6] G. Charpiat, O. Faugeras, and R. Keriven. Approximations of shape metrics and application to shape warping and shape statistics. Research Report 4820, INRIA, May 2003.

[7] C. Chefd'hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Regularizing flows for constrained matrix-valued images. *Journal of Mathematical Imaging and Vision*, 20(1-2):147–162, 2004.

[8] P.S. Eriksen. Geodesics connected with the fisher metric on the multivariate manifold. Technical Report 86-13, Institute of Electronic Systems, Aalborg University, 1986.

[9] P. T. Fletcher, C. Lu, and S. Joshi. Statistics of shape via principal geodesic analysis on lie groups. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 95–101, 2003.

[10] P.T. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Proc. Computer Vision Approaches to Medical Image Analysis*, Prague, May 2004.

[11] W. Förstner and B. Moonen. A metric for covariance matrices. Technical report, Stuttgart University, Dept. of Geodesy and Geoinformatics, 1999.

[12] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, X(IV):215–310, 1948.

[13] S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces.* Academic Press, 1978.

[14] H. Karcher. Riemannian centre of mass and mollifier smoothing. *Comm. Pure Appl. Math*, 30:509–541, 1977.

[15] C. Lenglet, M. Rousson, and R. Deriche. Segmentation of 3D probability density fields by surface evolution: Application to diffusion MRI. In *Proc. 7th Intl. Conf. on Medical Image Computing and Computer Assisted Intervention*, Saint-Malo, France, September 2004.

[16] C. Lenglet, M. Rousson, and R. Deriche. Toward segmentation of 3D probability density fields by surface evolution: Application to diffusion MRI. Research Report 5243, INRIA, June 2004.

[17] M. Moakher. Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications*, 24(1):1–16, 2002.

[18] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications (to appear)*, 2004.

[19] J.M. Oller and C.M. Cuadras. Rao's distance for negative multinomial distributions. *Sankhya: The Indian Journal of Stats.*, 47:75–83, 1985.

[20] X. Pennec. Probabilities and statistics on riemannian manifolds: A geometric approach. Research Report 5093, INRIA, January 2004.

[21] C.R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.

[22] M. Rousson, C. Lenglet, and R. Deriche. Level set and region based surface propagation for diffusion tensor mri segmentation. In *Computer Vision Approaches to Medical Image Analysis (CVAMIA) and Mathematical Methods in Biomedical Image Analysis (MMBIA) Workshop*, Prague, May 2004.

[23] L.T. Skovgaard. A riemannian geometry of the multivariate normal model. Technical Report 81/3, Statistical Research Unit, Danish Medical Research Council, Danish Social Science Research Council, 1981.

[24] L.T. Skovgaard. A riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, 11:211–233, 1984.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399